

Item Analysis for Korean-Working Memory Measures for Children (K-WMM-C)

Jiyun Han^a, Dongsun Yim^a, Hey-Eun Seok^b

^aDepartment of Communication Disorders, Ewha Womans University, Korea

^bDepartment of Psychology, Ewha Womans University, Korea

Correspondence: Dongsun Yim, PhD
Department of Communication Disorders, Ewha Womans University, 52 Ewhayecode-gil, Seodaemun-gu, Seoul 03760, Korea
Tel: +82-2-3277-6720
Fax: +82-2-3277-2122
E-mail: sunyim@ewha.ac.kr

Received: April 5, 2020
Revised: October 24, 2020
Accepted: November 9, 2020

This work was supported by the Ministry of Science and ICT of the Republic of Korea and the National Research Foundation of Korea (NRF-2019R1A2C1007488).

Objectives: This study performed item analyses of four working memory (WM) measures and constructed the validity of the WM measures. **Methods:** The present paper included two preliminary studies. The 1st preliminary study included data from three hundred twenty-nine children between 4-8 (NL 248, LI 81). Nonword repetition, matrix forward, matrix backward, and sentence repetition tasks were used. Item difficulty was analyzed using item response theory and classical test theory. In the 2nd preliminary studies, 84 children aged between 4-6 performed revised working memory tasks. Item difficulty and item discrimination parameters were analyzed. Lastly, construct validity was verified through the Pearson correlation between standardized vocabulary measures and four WM tasks. **Results:** Most items were appropriately placed in the order of difficulty, and some items with unacceptable difficulty levels were replaced, modified, or removed for the 2nd study. Most baseline items in forum WM tasks started with above 90% of the pass rate. The 2nd study results showed that item difficulty and discrimination parameters were adequate, but item parameters for few items were higher than expected, so that those items should be removed for the final version of tasks. The high level of concurrent validity for four WM tasks was demonstrated through significant correlations with receptive and expressive vocabulary scores. **Conclusion:** The present study systematically reviewed items in widely used WM measures. Based on the results, WM measures are expected to be used in the clinical setting. Further work is required to verify validity constantly.

Keywords: Working memory, Nonword repetition, Sentence repetition, Matrix, Item analysis, Item response theory

효과적인 언어중재를 계획하기 위해서는 아동의 현재 언어수준과 언어발달에 영향을 미치는 다양한 요인을 모두 고려한 다면적 평가가 중요하다. 다면적 평가에는 의미, 구문, 화용과 같은 전통적인 언어능력 평가와 작업기억과 같은 언어발달 기저의 정보처리기제 평가가 포함된다. 정보처리기제의 강점과 약점을 파악하는 것은 언어발달지체의 기전을 이해하고, 아동 개개인의 정보처리 용량 및 속도와 같은 특성을 고려하여 맞춤형 언어중재 프로그램을 계획할 수 있게 해준다. 또한 예후를 설명하기 위한 중요한 지표로 활용될 수 있다는 점에서 매우 중요하다 (Baddeley, Gathercole, & Papagno, 1998). 작업기억을 측정하기 위한 다양한 과제들의 임상적 효용성은 지속적으로 보고되고 있다(Alloway, Rajendran, & Archibald,

2009; Marton & Schwartz, 2003; Yang, Yim, Kim, & Han, 2013). 작업기억이란 짧은 시간동안 정보를 저장하고 조작할 수 있게 해주는 저장소이며 언어적 작업기억인 음운루프, 비언어적 작업기억인 시공간 잡기장 또는 시공간 스케치패드라는 하위시스템을 보유하고 있는 기억체계이다(Baddeley, 1997). 그 중 언어적 작업기억인 음운루프는 청각정보와 구어정보 처리 시 짧은 시간 동안 제한된 용량 내 저장하며, 어휘습득(Gathercole & Baddeley, 1989, 1993; Yang et al., 2013), 언어 이해(Just & Carpenter, 1992; Montgomery, 1995), 구문정보 처리(Adams & Gathercole, 1995; Han & Yim, 2016; King & Just, 1991), 읽기 이해(Boudreau, 2008; Daneman & Carpenter, 1980; Gathercole & Baddeley, 1993), 학업성취 및 문해력 습

득(Berninger, Nielsen, Abbott, Wijsman, & Raskind, 2008; Swanson & Berninger, 1995)과 같이 광범위하게 언어능력에 기여하는 것으로 알려져 있다. 언어적 작업기억 용량을 측정하기 위해 널리 사용되는 과제는 비단어 따라말하기 과제와 문장 따라말하기 과제이다. 비단어 따라말하기와 문장 따라말하기는 언어발달지체 아동에게 민감한 과제로(Archibald & Gathercole, 2006; Gathercole & Baddeley, 1990; Marton & Schwartz, 2003; Munson, Edwards, & Beckman, 2005; Yang et al., 2013) 국내외 다수의 연구에서도 임상적 효용성이 검토되었다(Oh & Yim, 2013; Yang & Yim, 2018; Yim, Yang, & Kim, 2015). 특히 Yim, Kim과 Yang (2016) 연구에서는 비단어 따라말하기와 문장 따라말하기 과제를 함께 사용했을 때 민감도 83.3%, 특이도 88.9%로 두 과제의 높은 언어발달지체 판별력을 보고하였다. 비단어 따라말하기 과제는 청각적으로 제시된 비단어를 듣고 따라말하는 과제로 생소한 음운조합의 청각정보를 처리하는 과정, 음운정보를 분석하고 저장한 뒤 회상하는 과정, 재산출을 위해 구어를 프로그래밍하고 실행하는 모든 일련의 과정을 포함한다(Gathercole, 2006). 문장 따라말하기는 문장을 듣고 기억했다가 재산출해야 하는 과제로 언어처리과정과 기억과정을 모두 포함하며 이원적인 측면의 작업기억의 효율성을 평가한다. 언어처리모델에 의하면 문장 따라말하기 과제를 수행하기 위해서는 청각정보 재인, 어휘 및 구문, 의미 표상이 필요하며 이후 재산출을 위한 구어 프로그래밍과 실행단계를 거쳐야 한다(MacDonald & Christiansen, 2002). 비단어와 달리 문장은 기억해야 하는 언어정보의 양이 많지만, 장기기억에 있는 구문 및 어휘 지식이 표상되어 문장 내 언어정보를 덩이 지어 기억하는데 도움이 되는 것으로 가정하기 때문에 기존의 언어지식에 의존하여 처리된다고 볼 수 있다. 현재 다양한 장애군과 목적에 따라 다양한 문장 따라말하기 패러다임 과제가 사용되고 있으며(Hwang et al., 2016; Yoon, 2015) 사용하는 과제에 따라 아동의 수행이 다르게 평가될 수 있기 때문에 임상에서 쉽게 이러한 처리과제를 사용하는 데에는 어려움이 따른다.

또한 시공간 정보와 같이 비언어정보를 처리하는 작업기억이 있다. 시공간 스케치패드라고 불리는 작업기억 용량은 시, 공간적인 정보를 짧은 시간동안 저장하는 곳으로서, 시각정보의 패턴이나 공간적 위치와 제시순서를 기억하는데 관여한다(Wilson, Scott, & Power, 1987). 일반인지처리이론(*general cognitive processing theory*)의 관점에서 아동의 낮은 정보처리 능력 및 느린 처리속도는 영역 일반적으로 나타나며, 언어적, 비언어적 정보처리에 모두 영향을 줄 수 있다. 현재 이러한 비언어적 작업기억을 평가하기 위한 과제로 시각정보를 기억했다가 산출해야 하는 매트릭스 과제가 널리 사용되고 있다. 정보 저장단계를 측정하기 위해 매트릭스-정

순회상 과제를 실시하고, 저장과 처리의 이원적인 면을 측정하고자 매트릭스-역순회상 과제를 실시한다. 순차적으로 본 것을 그대로 산출하게 할 경우 단순히 공간적 정보를 저장하는 과정까지 포함되지만, 반대로 제시된 순서를 거꾸로 회상해야 하는 경우에는 공간 정보를 저장했다가 거꾸로 정보를 회상하는 과정에서 작업기억의 중앙집행기에서 주의비용을 더 할애하여 정보를 조작해야 하는 2차 처리과정까지 포함한다(Monaco, Costa, Caltagirone, & Carlesimo, 2012). 따라서 아동의 매트릭스 정순, 역순 과제를 통해 아동의 정보처리적 특성을 단계 별로 살펴볼 수 있다.

그러나 이러한 작업기억 측정과제들은 임상 현장에서 평가목적으로 사용하는 데에는 한계가 있다. 첫째, 아동의 연령, 언어능력에 따라 작업기억 수행이 유의하게 달라지는데, 대부분의 과제는 모든 아동 대상자에게 처음부터 끝까지 동일하게 실시한다. 특히 만 4세 이하의 아동은 문항 수가 많거나 복잡한 단계를 포함하는 작업기억 측정과제 실시 시 오래 착석하여 집중하는 데 어려움이 있기 때문에 과제의 전 문항을 실시하는 것은 비효율적일 수 있다. 특히, 작업기억 측정과제 특성 상 주어진 자극을 순서대로 혹은 역순으로 기억해야 하는데 아동의 연령과 발달수준을 고려하지 않고 많은 수의 문항으로 구성된 과제를 실시할 경우 과제 상황이 익숙하지 않은 취학 전 아동에게는 거부감을 일으키는 원인이 될 수 있다. 그리고 언어발달지체 아동의 경우, 제한된 정보처리 용량으로 인해 상위/고도의/높은 통합적 인지능력을 필요로 하는 작업기억 측정과제 수행 시, 집중력 있게 모든 과제를 수행하는 데에 어려움이 따를 수 있다(Archibald & Gathercole, 2006). 따라서 연령이 낮은 아동의 경우 연령이 높은 아동에 비해 주의 집중 시간이 짧거나, 충동적일 수 있으며, 이러한 행동적 특성이 평가에 영향을 미칠 수 있다는 점을 감안하여 작업기억 측정과제 내 단계 및 연령 별 수행기준을 마련하여 측정의 효율성과 정확성을 높일 필요가 있다(Leonard et al., 2007).

둘째, 연구에서 사용하는 작업기억 측정과제가 다양하다. 그리고 다양한 과제를 통해 도출된 결과는 연구마다 다르다. 따라서 임상에서 사용하기에는 적절한 도구 선택부터, 평가 결과를 해석하는 데까지 다소 어려움이 따른다. 이에, 과제의 문항별 난이도와 같은 문항양호도 검사를 통해 타당하게 구성된 작업기억 측정도구가 필요하다.

셋째, 과제가 구인을 정확히 측정할 수 있는 문항으로 구성되어 있는지 문항양호도를 살펴야 한다. 이때 검사이론(*test theory*)의 고전검사이론(*classical test theory, CTT*)과 문항반응이론(*item response theory, IRT*)에 의거하여 과제별 문항양호도를 검증할 수 있다(Lawley, 1943; Sung, 2016). 문항양호도는 문항난이도와 문항

변별도 지수를 추정하여 확인할 수 있다. 우선 고전검사이론에 의한 문항난이도는 응답한 피험자 중 정답을 맞힐 피험자의 비율이다. 반면 문항반응이론의 난이도는 개별 문항의 문항특성곡선(item characteristic curves, ICC)을 바탕으로 답을 맞출 확률이 0.5에 대응하는 능력수준의 값을 추정하는 방식을 따른다. 문항변별도는 피험자의 능력을 본 문항이 어느 정도 변별해 내는지 확인하는 지수이다.

고전검사이론에서는 응답자 표본의 특성에 따라 난이도와 변별도 같은 문항특성이 다르게 추정된다는 제한점이 있다(Hambleton & Swaminathan, 1985). 예를 들어 같은 문항에 대해 일반 아동은 100명 중 70명이 맞추고, 언어발달지체 아동은 100명 중 30명이 맞춘다면 해당 문항의 난이도는 각각 .7과 .3으로 피험자 특성에 따라 나뉘게 되어 동일한 문항이 피험자 집단에 따라 난이도가 다르게 산출되는 것이다. 이 경우, 쉬운 검사에서 피험자 능력이 과대 추정되고, 어려운 검사에서 피험자 능력이 과소 추정될 수 있다. 반면에 문항반응이론은 응답자 표본의 특성에 영향을 받지 않고 아동의 검사 결과에서 나타난 성취와 잠재된 작업기억 능력과의 관계를 추정하는 이론으로, 문항이 불변하는 고유한 속성을 문항특성곡선에 의해 분석하는 검사 이론이다. 문항반응이론에 의한 분석이란 이론적으로 일반 아동이나 언어발달지체 아동과 같이 집단이 바뀌어도 문항난이도에는 변화가 없다는 것을 가정한다. 문항반응이론은 고전검사이론의 한계점을 극복하는 대안적인 방법으로 최근 검사의 양호도 검증 연구에서 주목받고 있다(Embretson & Reise, 2000; Sung, 2002; 2016). 그러나 문항반응이론에 근거하여 문항양호도를 안정적으로 추정하기 위해서는 자료가 문항반응이론을 적용하기 적합한 자료인지 검토해야 한다. 표본의 크기가 문항당 1,000명 정도로 커야 하며, 자료가 문항반응이론의 기본 가정인 일차원성 가정, 문항독립성 가정을 만족해야 한다. 일차원성 가정이란 하나의 검사는 하나의 특성만을 측정한다는 것을 의미하며(하나의 검사가 측정하는 영역이 여러 개일 수 없음), 문항독립성이란 어느 한 문항이 다른 문항을 수행하는데 영향을 미쳐서는 안된다는 것을 의미한다. 표본의 크기가 매우 커야한다는 제한점이 있다.

문항양호도 검증을 위해서는 기존의 고전검사이론보다 이론적으로 여러 장점을 보유한 문항반응이론도 함께 적용하여 아동에게 더 적절한 방법으로 작업기억 과제들의 타당도의 증거를 제시할 수 있는 방안을 모색해야 한다.

종합하면, 적은 수의 문항으로도 아동의 정보처리 특성을 쉽게 평가할 수 있는 통일된 작업기억 측정도구가 필요하다. 이에 본 연구에서는 두 번에 걸친 예비검사를 통해 아동의 작업기억 척도의 문항들을 분석하였다. 첫 번째 1차 예비검사 단계에서는 선행연구

에서 발표한 자료를 취합하여 연령, 언어 집단, 성별에 따라 분석하였으며, 분석내용을 토대로 문항난이도와 문항변별도를 산출하여 검사항목을 수정 및 보완하였다. 이때 문항양호도를 살펴보기 위해 고전검사이론과 문항반응이론에 근거하여 점수를 도출하고 비교하여, 언어발달지체 아동 변별을 가능하게 하였다. 작업기억 측정과제의 타당도를 검증함에 있어 이러한 문항난이도와 문항변별도를 제시하는 것은 과제 수행시간을 줄이고, 연령별로 적합한 문항을 제시할 수 있는 근거가 되기 때문에 이후 임상 현장에서의 활용도를 높일 수 있을 것으로 기대한다. 이후 2차 예비검사를 통해 수정된 과제들의 문항 양호도를 재검토 하여 아동의 작업기억 발달 수준 평가 과제의 정확성을 높였다.

본 연구는 언어발달지체 및 일반 아동을 포함한 총 413명의 아동을 대상으로 비언어적, 언어적 작업기억 용량을 측정함에 있어 고전검사이론 및 문항반응이론에 의거하여 문항난이도와 문항변별도를 산출하여 문항양호도를 검증하였다. 두 차례의 예비검사 단계를 거쳐 문항들을 재정비하였고 문항양호도의 변화 양상을 살펴보았다. 본 연구를 수행하기 위한 구체적인 연구질문은 다음과 같다.

1. 1차 예비검사를 통해 실시한 작업기억 측정과제(비단어 따라 말하기, 문장 따라말하기, 매트릭스 정순, 역순)의 문항난이도와 문항변별도 지수는 어떠한가?
2. 2차 예비검사에서 선정한 문항난이도와 문항변별도 지수는 어떠한가?
3. 아동용 작업기억 과제의 공인타당도는 어떠한가?

연구방법

연구대상

1차 예비검사 단계에서 본 연구에 포함된 자료는 2014년부터 2016년에 걸쳐 수집된 자료로, 서울, 경기, 인천 지역에 거주하는 만 4-8세 아동 329명(남자 165명, 여자 164명)이며 이 중 일반 아동은 248명, 언어발달지체 아동은 81명이었다(문장 따라말하기 제외, 문장 따라말하기의 경우 일반 아동 168명, 언어발달지체 아동 40명을 대상으로 함). 본 자료는 Kim과 Yim (2015), Yang 등(2013), Yang, Yim과 Bae (2015), Yim 등(2016)에 사용된 자료를 2차로 사용하였다. 2차 예비검사 단계에서는 서울, 경기 지역에 거주하는 만 4-6세 아동 84명(남자 45명, 여자 39명)을 대상으로 하였으며 이 중 일반 아동은 56명, 언어발달지체 아동은 28명이었다. 이로써 1차, 2차 시기에 누적된 피험자 수는 총 413명이었다(문장따라말하기 제외).

연구에 포함된 언어발달지체 아동은 1) 카우프만 아동용 지능검사(K-ABC; Moon & Byun, 2003)의 동작성 지능지수 85 (-1 SD) 이

상, 2) 수용 및 표현어휘력 검사(REVT; Kim, Hong, Kim, Jang, & Lee, 2009) 수용 또는 표현어휘력 중 한 개 이상의 영역에서 10%ile 미만 또는 취학전 아동의 수용언어 및 표현언어 발달 척도(PRES; Kim, Sung, & Lee, 2003) 검사 결과 수용 또는 표현언어가 10%ile 미만이며 3) 부모보고에 의해 감각, 정서, 행동, 신경학적 문제가 없는 아동으로 하였다. 일반 아동은 1) K-ABC 동작성 지능지수 85 (-1 SD) 이상, 2) REVT 수용 및 표현어휘 모두 10%ile 이상, 3) PRES 수용 및 표현언어 모두 10%ile 이상, 4) 감각, 정서 및 행동, 신경학적 결함이 없는 아동으로 하였다. 선행연구를 통해 과제의 타당도 및 신뢰도의 증거를 확보하였으나 과제들이 측정하고자 하는 구인을 잘 측정하고, 더불어 현 단계에서 언어발달지체 아동과 일반 아동을 구분하기에 적합한 과제인지 검토하기 위해 두 집단의 아동을 모두 대상으로 하였다. 대상자 특성은 Table 1에 제시하였고, 집단 내 연령 별 대상자 수는 Appendix 1에 제시하였다.

연구도구

비단어 따라말하기 과제는 2-6음절의 비단어를 들려주고 아동이 재산출하도록 하는 과제이다. 예비검사 단계에서는 각 단계 별 4개씩 총 20문항으로 구성된 과제를 사용하였다. 비단어 따라말하기 수행은 음절길이(Dollaghan & Campbell, 1998), 단어유사성 정도(Munson, Kurtz, & Windsor, 2005)에 영향을 받는다. 이에 음절 길이를 2음절부터 6음절로 하여 문항의 난이도를 높였으며, 단어 유사성이 높으면 어휘지식에 의존하여 과제를 수행할 수 있기 때문에 단어유사성이 낮은 비단어로 난이도를 조정하였다. 음절구조는 기본적으로 열린 음절로 구성하여 2음절 CVCV, 3음절 CVCVCV, 4음절 CVCVCVCV, 5음절 CVCVCVCVCV, 6음절 CVCVCVCVCVCV로 하였으나 일부 음절에는 CVC 또는 VC도 포함하였다. 자음의 경우 습득시기가 늦은 마찰음 /s, ʃ/와 말소리가 약하게 발화될 경우 지각이 어려울 수 있는 성문음 /h/는 제외하였고, 모음은 단모음(/i, u, e, o/) 제외)으로 구성하였다. 비단어는 Muse

(2003)가 제시한 바에 따라 실제 단어의 첫 음소를 조음방법과 조음위치를 변형한 자음으로 바꾼 뒤, 각 음절의 초성 자음의 위치를 바꾸는 방법으로 제작하였으며, 모음과 종성자음은 바꾸지 않았다. 점수 산출은 낱말 및 음절 단위로 산출하며 정반응은 1점, 오반응은 0점으로 채점하였다.

매트릭스 정순, 역순 과제는 3×3 배열의 정사각형을 9개가 순차적으로 한 개씩 점등되는 것을 기억했다가 회상해야 하는 과제이다. 점등되는 양은 2개부터 5개로 점차 증가한다. 비단어 따라말하기와 문장 따라말하기와 같이 음운기억에 의존하여 과제를 수행하는 방법이 아닌 음운정보를 배제하고 아동이 순차적 또는 역순차적으로 제시된 정보를 회상할 수 있는지를 통해 아동의 작업기억 용량을 측정하기 위해 고안된 과제이다. 기억용량은 자극의 양에 반비례하므로, 제시하는 자극을 2개에서 5개까지 점차 늘리는 방법으로 과제의 난이도를 조정하였다. 자극은 1개씩 0.5초 동안 점등되고, 0.5초 간격으로 제시되었다. 매트릭스 역순 과제는 정순 과제와 겹치지 않도록 시각 도안을 변경하여 제작하였고, 실시방법은 동일하나 아동이 정보를 회상해야 할 때 마지막에 본 자극부터 역순으로 회상하도록 하였다. 예비검사에서 사용한 과제는 2-3개 제시되는 단계는 4문항씩, 4-5개 제시되는 단계에서는 5문항씩 제시되어 총 18문항으로 구성된 과제를 사용하였다. 점수 산출을 위해 아동의 수행을 토대로 문항 별 0점, 1점으로 정오반응을 기록한다.

문장 따라말하기 과제는 3-5어절의 문장을 들려주고 아동이 재산출하도록 하는 과제이다. 문장 따라말하기 수행은 문장길이와 문장구조에 영향을 받기 때문에 문장길이를 3어절과 5어절 문장으로 구성하였고, 문장구조는 3세 아동에게서 가장 빈번하게 나타나는 구문구조인 단문, 접속복문(시간, 이유, 조건: Lee, Kwon, & Kim, 1979) 및 내포복문(명사절, 인용절, 관형절: Kim, 1997)을 포함하여 난이도를 조정하였다. 3어절 문장은 단문으로만 구성하였으며 목적어-부사어-서술어, 주어-목적어-서술어, 주어-부사어-서술어의 구문구조로 각각 3문항씩 총 9문항을 구성하였다. 5어절

Table 1. Participants' characteristics

Characteristic	1st preliminary study			2nd preliminary study		
	NL (N=248)	LI (N=81)	t	NL (N=56)	LI (N=28)	t
Age (mo)	67.83 (10.04)	69.95 (11.44)	-1.592	62.15 (9.66)	61.34 (11.10)	-.371
Performance IQ ^a	110.65 (22.44)	104.21 (12.72)	2.456*	107.91 (13.86)	96.03 (11.32)	-4.189**
Receptive vocabulary ^b	66.77 (15.34)	47.33 (15.26)	9.908**	60.17 (14.10)	41.23 (19.76)	-4.968**
Expressive vocabulary ^b	72.22 (13.67)	57.72 (14.58)	7.041**	65.88 (12.26)	51.29 (18.31)	-4.190**

Values are presented as mean (SD).

NL=children with normal language development; LI=children with language impairment.

^aKaufman assessment battery for children (Moon & Byun, 2003), ^bReceptive & expressive vocabulary test (Kim, Hong, Kim, Jang, & Lee, 2009).

*p<.05, **p<.01.

문장은, 단문, 접속복문, 내포복문으로 구성하였는데, 단문은 주어-관형어-부사어-목적어-서술어, 주어-부사어-관형어-목적어-서술어, 주어-부사어-목적어-부사어-서술어로 구성된 문장을 각각 3문항씩 총 9문항 제작하였고, 접속복문은 시간절(주어-목적어-서술어)-목적어-서술어, 이유절(주어-서술어)-목적어-부사어-서술어, 조건절(주어-서술어)-주어-목적어-서술어로 구성된 문장을 각각 3문항씩 9문항 구성하였다. 마지막으로 내포복문은 주어-명사절(목적어-서술어)-서술어, 주어-관형절(목적어-서술어)-목적어-서술어, 주어-인용절(목적어-부사어-서술어)-서술어 구조로 각 구문구조 당 3문항씩 총 9문항을 제작하였다. 모든 문장에 포함된 낱말은 3-5세 아동의 회화에 등장하는 어휘로 구성하였다(Lee, Jung, Lee, & Lee, 1972). 예비검사 단계에서는 3어절 단문 9문항, 5어절 단문 9문항, 5어절 접속복문 9문항, 5어절 내포복문 8문항으로 총 34문항으로 구성된 과제를 사용하였다. 점수는 정반응은 1점 오반응은 0점으로 채점하여 산출하였다.

2차 예비검사 단계에서는 1차 예비검사 결과를 반영하여 최종 선정된 문항으로 재구성된 과제를 사용하였다. 최종 문항 선정을 위해 문항난이도가 낮은 문항은 삭제하였고, 난이도가 유사한 문항이 중복적으로 제시되지 않게 각 단계를 대표할 수 있는 문항만 포함하였으며, 문항 재배치를 통해 문항이 올라갈수록 난이도가

높아지도록 과제를 재구성하였다. 이에, 비단어 따라말하기 과제는 각 단계 별 3문항씩 제시하여 총 20문항에서 15문항으로 줄였고, 매트릭스 정순, 역순 과제는 점등 순서 및 점등 패턴이 앞 문항과 비슷하여 난이도가 비슷하게 나온 문항을 수정하고 단계가 높지만 난이도가 전 단계보다 낮게 나온 문항은 점등 순서를 조정하여 난이도를 높였다. 총 문항 수는 그대로 18문항으로 하였다. 마지막으로 문장 따라말하기 과제는 아동의 집중력을 고려하여 총 34문항에서 각 단계 별 3문항씩 총 12문항으로 조정하였다. 조정된 문항의 세부내용은 Appendix 2, 3, 4에 제시하였고, 1차, 2차 예비검사 구성 내용은 Table 2에 제시하였다.

자료분석

검사 및 문항의 양호도를 분석하기 위해 고전검사이론(CTT)과 문항반응이론(IRT)을 통해 양호도 지수인 난이도, 변별도를 산출하였다. 고전검사이론에 의한 난이도는 응답 피험자 중 정답을 맞힌 피험자의 비율을 산출하고 변별도는 검사 총점과 문항 점수 간의 양수 관계수로 추정하지만, 문항반응이론의 난이도는 개별 문항의 문항특성곡선(ICC)을 바탕으로 답을 맞힐 확률이 .5에 대응하는 능력수준의 값으로 정하여 난이도를 정한다. 그리고 변별도는 난이도에 해당하는 능력수준에서 ICC의 기울기로 추정한다.

Table 2. Preliminary versions of NWR, Matrix forward, Matrix backward, and SR tasks

Task	1st preliminary study			2nd preliminary study		
	Item number	Number of items	Total items	Item number	Number of items	Total items
Nonword repetition			20			15
2 syllables	1-4	4		1-3	3	
3 syllables	5-8	4		4-6	3	
4 syllables	9-12	4		7-9	3	
5 syllables	13-16	4		10-12	3	
6 syllables	17-20	4		13-15	3	
Matrix forward			18			18
2 blocks	1-4	4		1-4	4	
3 blocks	5-8	4		5-8	4	
4 blocks	9-13	5		9-13	5	
5 blocks	14-18	5		14-18	5	
Matrix backward			18			18
2 blocks	1-4	4		1-4	4	
3 blocks	5-8	4		5-8	4	
4 blocks	9-13	5		9-13	5	
5 blocks	14-18	5		14-18	5	
Sentence repetition			34			12
Simple sentences: 3 words	1-9	9		1-3	3	
Simple sentences: 5 words	10-18	9		4-6	3	
Conjoined sentences: 5 words	19-27	9		7-9	3	
Embedded sentences: 5 words	1-8	8		10-12	3	

본 연구의 문항난이도와 변별도는 R. 3.5.3 (Team, 2013)을 이용하여 산출하였다. 고전검사이론에 기반한 문항난이도와 문항변별도는 psychometric (Schumacker, 2019) 패키지의 item. Exam 함수, 문항반응이론에 기반한 문항난이도와 문항변별도는 irtoys (Partchey 외, 2017) 패키지의 est 함수를 이용하여 2모수 모형을 적용하여 문항모수치를 추정하였다. 문항난이도 그리고 SPSS ver 26을 사용하여 공인타당도를 산출하였다.

연구결과

인구통계학적 분포

1차 예비검사에 사용된 표본의 구성은 Table 2와 같다. 1차 예비검사에 포함된 자료는 일반 아동 248명, 언어발달지체 아동 81명의 자료였고, 문장 따라말하기의 경우 일반 아동 168명, 언어발달지체 아동 40명의 자료를 사용하였다. 각 과제 별 아동이 총점 0점을 맞거나 지시문을 이해하지 못해 과제를 수행하지 못한 '불능'의 경우 분석에서 제외되었다. 과제 별 '불능'으로 처리된 인원은 비단어 따라말하기 과제에서는 일반아동 5명, 언어발달지체 아동 10명, 매트릭스 정순 과제에서 일반아동 5명, 언어발달지체 아동 14명, 매트릭스 역순 과제에서 일반아동 37명, 언어발달지체 아동 27명이었다. 문장 따라말하기 과제의 경우 전체 참여자가 전원이 불능 없이 실시하였다. 과제 별 아동의 참여율은 Table 3에 제시하였다. 남녀 성별은 각각 165명, 164명으로 비슷한 비율이었으며, 만 4세 참여인원은 65명, 만 5세는 135명, 만 6세는 109명으로 많았으나, 만 7세는 11

명, 만 8세는 9명으로 수가 매우 적었다. 신뢰로운 결과를 얻기 위해서 충분한 표본 크기가 확보되어야 하기 때문에 2차 예비검사에서는 수가 매우 적었던 연령군을 제외하여 수집하였다.

기초통계분석

4개의 작업기억 측정과제 총점과 각 과제의 하위영역의 총점에 대한 평균과 표준편차를 산출하였다. Table 4에 제시한 바와 같이, 전반적으로 각 과제의 하위영역들의 총점 평균은 각 검사의 뒤쪽으로 배치될수록 낮게 나타났다. 그러나, 문장 따라말하기 과제의 경우 5어절 접속복문(5-word conjoined sentences)의 총점 평균이 5어절 단문(5-word simple sentences)보다 뒤에 배치되었음에도 총점 평균이 더 높게 나타났다. 아동이 접속복문보다 단문을 더 어려워하는 것으로 보이며, 2차 예비검사 시 문항재배치를 고려해야 하는 것으로 보인다.

성별 및 집단에 따른 과제별 수행도를 Table 4에 함께 제시하였다. 여아와 남아의 과제별 수행은 전반적으로 유사하였다. 그리고 일반 아동과 언어발달지체 아동의 과제별 평균은 전반적으로 일반 아동의 평균이 높게 나타났으며, 각 과제의 첫 번째 단계에서는 수행이 거의 같지만 단계가 올라갈수록 집단별 평균 차이가 커지는 것으로 나타났다. 그리고 문장 따라말하기 과제의 5어절 단문(5-word simple sentences) 평균에서 일반 아동과 언어발달지체 아동 간의 수행차가 크게 나타났다.

이어서 연령별 수행을 Table 5에 제시하였다. 전반적으로 연령이 높아지며 평균 점수가 높아지는 양상이 나타났다.

Table 3. Participants' characteristics by working memory tasks

Group	N	NWR	Matrix (F)	Matrix (B)	SR
LI	81 (24.67)	71 (22.61)	67 (21.61)	54 (20.45)	40 (19.23)
NL	248 (75.38)	243 (77.39)	243 (78.39)	210 (79.55)	168 (80.77)
Sex					
Boy	165 (50.15)	157 (50.00)	156 (50.32)	132 (50.00)	102 (49.04)
Girl	164 (49.85)	157 (50.00)	154 (49.68)	132 (50.00)	106 (50.96)
Age					
4	65 (19.76)	61 (19.43)	60 (19.35)	21 (7.95)	- (-)
5	135 (41.03)	132 (42.04)	130 (41.94)	127 (48.11)	118 (56.73)
6	109 (33.13)	103 (32.80)	100 (32.26)	97 (36.74)	84 (40.38)
7	11 (3.34)	11 (3.50)	11 (3.55)	10 (3.79)	6 (2.88)
8	9 (2.74)	7 (2.23)	9 (2.90)	9 (3.41)	- (-)
Total	329 (100)	314 (100)	310 (100)	264 (100)	208 (100)

Values are presented as number (Percentage).

LI=children with language impairment; NL=children with normal language development; NWR=nonword repetition; Matrix(f)=Matrix forward; Matrix(B)=Matrix backward; SR=sentence repetition.

예비검사용 작업기억 측정과제의 문항분석

문항난이도는 지수로 산출되며, Baker (2001)는 문항반응이론에서의 문항난이도 해석 근거로 $-\infty \sim -2.0$ 매우 쉬움, $-2.0 \sim -1.5$ 쉬움, $-1.5 \sim +1.5$ 중간, $+1.5 \sim +2.0$ 어려움, $+2.0 \sim \infty$ 매우 어려움을 제안하였다. 그리고 Sung (1991)에서는 고전검사이론에서의 문항난이도 해석 근거로 $0 \sim 1.0$ 매우 쉬움, $1.0 \sim 2.0$ 쉬움, $2.0 \sim 3.0$ 중간, $3.0 \sim 4.0$ 어려움, $4.0 \sim \infty$ 매우 어려움을 제안하였다. 문항변별도는 지수로 산출되며, Baker (2001)는 문항반응이론에서의 문항변별도 해석근거로 $1.7 \sim \infty$ 매우 높음, $1.35 \sim 1.70$ 높음, $0.65 \sim 1.35$ 중간, $0.35 \sim 0.65$ 낮음, $0 \sim 0.35$ 매우낮음, $0 \sim \infty$ 부적변별도를 제안하였고, Ebel (1965)은 고전검사이론에서의 문항변별도 해석근거로 $0.4 \sim 1.0$ 매우 높음/높음, $0.3 \sim 0.4$ 중간, $0.2 \sim 0.3$ 낮음, $0 \sim 0.2$ 매우 낮음, $-1.0 \sim 0$ 부적변별도를 제안하였다. 문항난이도와 문항변별도의 언어적 해석은 Table 6에 제시하였다.

비단어 따라말하기(NWR)

Table 4. Mean scores and SD for NWR, Matrix (F), Matrix (B), and SR

	Total			Boys			Girls			NL			LI		
	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD
Nonword repetition															
2 syllables	314	3.45	0.71	157	3.48	0.75	157	3.43	0.68	243	3.45	0.72	71	3.45	0.69
3 syllables	314	3.09	0.88	157	3.01	0.85	157	3.17	0.91	243	3.14	0.83	71	2.92	1.02
4 syllables	314	2.80	1.10	157	2.66	1.14	157	2.94	1.04	243	2.88	1.07	71	2.49	1.16
5 syllables	314	2.01	1.21	157	1.89	1.24	157	2.14	1.18	243	2.07	1.21	71	1.80	1.20
6 syllables	314	1.36	1.20	157	1.27	1.28	157	1.45	1.12	243	1.42	1.18	71	1.14	1.26
NWR total	314	12.71	3.54	157	12.31	3.67	157	13.11	3.38	243	12.98	3.48	71	11.80	3.64
Matrix forward															
2 blocks	310	3.55	0.76	156	3.44	0.77	154	3.65	0.73	243	3.57	0.72	67	3.46	0.88
3 blocks	310	2.71	1.28	156	2.71	1.30	154	2.71	1.27	243	2.72	1.27	67	2.66	1.33
4 blocks	310	2.00	1.56	156	1.96	1.55	154	2.05	1.57	243	2.02	1.55	67	1.91	1.57
5 blocks	310	0.96	1.38	156	0.96	1.42	154	0.97	1.34	243	0.94	1.36	67	1.03	1.47
Matrix (F) total	310	9.22	4.02	156	9.06	4.09	154	9.38	3.95	243	9.26	3.96	67	9.06	4.25
Matrix backward															
2 blocks	264	3.25	1.05	132	3.18	1.14	132	3.33	0.95	210	3.27	1.00	54	3.20	1.23
3 blocks	264	2.42	1.33	132	2.37	1.39	132	2.48	1.27	210	2.43	1.32	54	2.39	1.38
4 blocks	264	1.45	1.49	132	1.46	1.52	132	1.45	1.47	210	1.48	1.45	54	1.37	1.65
5 blocks	264	0.72	1.20	132	0.74	1.19	132	0.70	1.21	210	0.74	1.21	54	0.63	1.14
Matrix (B) total	264	7.85	3.98	132	7.76	4.10	132	7.95	3.88	210	7.92	3.91	54	7.59	4.30
Sentence repetition															
3 words SS total	208	8.73	0.60	102	8.70	0.54	106	8.76	0.66	168	8.80	0.46	40	8.45	0.96
5 words SS total	208	6.58	2.33	102	6.25	2.46	106	6.89	2.17	168	7.07	1.78	40	4.53	3.15
5 words CS total	208	7.22	1.93	102	7.02	2.02	106	7.41	1.82	168	7.56	1.55	40	5.78	2.60
5 words ES total	208	5.46	2.20	102	5.45	2.13	106	5.46	2.28	168	5.83	1.95	40	3.90	2.52
SR total	208	27.98	5.91	102	27.42	5.97	106	28.52	5.83	168	29.25	4.50	40	22.65	7.91

SS=simple sentence; CS=conjoined sentence; ES=embedded sentence; NL=children with normal language development; LI=children with language impairment.

앞서 언급한 기준에 근거하였을 때, 문항반응이론에 근거한 문항난이도 지수는 2음절 매우 쉬움, 3음절 매우 쉬움-쉬움, 4음절 쉬움-중간, 5음절, 6음절 중간-어려움으로 산출되었으며, 고전검사이론에 근거한 난이도는 2음절 매우 쉬움, 3음절 쉬움, 4음절 쉬움-중간, 5음절 중간-어려움, 6음절 어려움으로 나타났다. 문항난이도의 경우 문항반응이론 보다 고전검사이론에 근거한 문항난이도 지수가 매우 쉬움부터 어려움까지 고르게 퍼져있는 것으로 나타났다.

문항변별도 지수의 경우, 선행연구에 근거하여 결과를 해석하였을 때, 2음절은 낮음, 3음절은 낮음-중간, 4음절 중간, 5음절 중간-높음, 6음절 중간 수준으로 나타났다. 고전검사이론에 근거한 문항변별도는, 2음절은 낮음, 3음절과 4음절은 낮음-중간, 5음절은 중간-높음, 6음절은 높음으로 나타나 문항반응이론에 근거한 문항변별도와 달리 2음절부터 6음절까지 음절이 높아지면서 변별도가 점차 높아지는 양상으로 산출되었다. 문항반응이론과 고전검사이론 모형에 근거한 문항양호도 지수와 문항 별 순위를 Table 7에 제시

하였다.

매트릭스 정순

앞서 언급한 기준에 근거하였을 때, 문항반응이론에 근거한 문항난이도 지수는 2개 블록이 점등되는 조건은 매우 쉬움-쉬움, 3개 블록 조건에서는 쉬움-중간, 4개 블록 조건에서는 중간-어려움, 5개 블록 조건에서는 어려움으로 추정되었고, 고전검사이론에 근거한 난이도는 2개 블록 조건은 매우 쉬움-쉬움, 3개 블록 조건은 쉬움, 4개 블록 조건은 중간-어려움, 5개 블록 조건에서는 어려움-매우 어려움으로 나타났다. 대체적으로 두 검사이론에 근거한 문항난이도 지수가 유사하게 추정되었으나, 문항반응이론에 근거한 문항난이도 중 4개 블록이 점등되는 조건의 4번 문항이 난이도가 -.269로 나타나 주변 문항에 비하여 난이도가 쉽고, 문항배치가 적절하지 검토할 필요가 있는 것으로 나타났다.

문항변별도를 분석한 결과, 문항반응이론에 근거한 문항변별도

Table 5. Mean scores and SD for NWR, Matrix (F), Matrix (B), and SR by ages

Task	4			5			6			7			8		
	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD
Nonword repetition															
2 syllables	61	3.26	0.81	132	3.44	0.70	103	3.58	0.65	11	3.64	0.67	7	3.14	0.69
3 syllables	61	2.75	0.99	132	3.14	0.87	103	3.24	0.80	11	3.27	0.47	7	2.57	0.98
4 syllables	61	2.03	1.18	132	2.80	1.05	103	3.14	0.94	11	3.36	0.67	7	3.57	0.53
5 syllables	61	1.21	1.07	132	2.00	1.14	103	2.39	1.17	11	2.82	1.08	7	2.43	1.13
6 syllables	61	0.95	1.06	132	1.22	1.15	103	1.68	1.22	11	2.09	1.45	7	1.71	1.38
NWR total	61	10.21	3.63	132	12.59	3.29	103	14.03	3.09	11	15.18	2.14	7	13.43	3.55
Matrix forward															
2 blocks	60	3.12	1.01	130	3.55	0.74	100	3.71	0.54	11	3.91	0.30	9	4.00	0.00
3 blocks	60	1.68	1.30	130	2.56	1.21	100	3.36	0.92	11	3.18	0.98	9	3.89	0.33
4 blocks	60	0.85	1.09	130	1.70	1.29	100	2.84	1.50	11	2.55	1.69	9	4.00	1.32
5 blocks	60	0.08	0.38	130	0.57	1.01	100	1.71	1.51	11	1.64	1.69	9	3.33	1.50
Matrix (F) total	60	5.73	2.97	130	8.38	3.20	100	11.62	3.50	11	11.27	3.50	9	15.22	2.33
Matrix backward															
2 blocks	21	1.76	1.55	127	3.23	0.98	97	3.49	0.78	10	3.80	0.42	9	3.89	0.33
3 blocks	21	1.06	1.24	127	2.17	1.27	97	2.87	1.18	10	3.20	0.92	9	3.56	0.73
4 blocks	21	0.33	0.58	127	1.10	1.25	97	1.84	1.53	10	2.80	1.69	9	3.44	1.59
5 blocks	21	0.10	0.30	127	0.38	0.83	97	1.11	1.44	10	1.40	1.07	9	2.00	1.66
Matrix (B) total	21	3.24	3.02	127	6.88	3.15	97	9.31	3.88	10	11.20	3.08	9	12.89	3.52
Sentence repetition															
3 words SS total	0	.	.	118	8.72	0.65	84	8.75	0.53	6	8.67	0.52	0	.	.
5 words SS total	0	.	.	118	6.36	2.38	84	6.95	2.21	6	5.50	2.59	0	.	.
5 words CS total	0	.	.	118	7.01	1.91	84	7.49	1.96	6	7.50	1.38	0	.	.
5 words ES total	0	.	.	118	5.19	2.24	84	5.83	2.16	6	5.50	1.64	0	.	.
SR total	0	.	.	118	27.28	6.01	84	29.02	5.75	6	27.17	4.71	0	.	.

SS=simple sentence; CS=conjoined sentence; ES=embedded sentence.

Table 6. Interpretation for item difficulty and item discrimination indices by IRT and CTT

Interpretation	Item difficulty		Item discrimination		
	IRT	CTT	Interpretation	IRT	CTT
Very easy	-∞~-2.0	0-1.0	Very high	1.7-∞	.4-1.0
Easy	-2.0~-5	.6-.8	High	1.35-1.70	.3-.4
Moderate	-.5~+.5	.4-.6	Moderate	.65-1.35	.2-.3
Challenging	+.5~+2.0	.2-.4	Low	.35-.65	.1-.2
Very challenging	+2.0~∞	0-.2	Very low	0-.35	.1-0

IRT = item response theory; CTT = classical test theory.

지수는 2개 블록조건부터 5개 블록 조건까지 대체적으로 중간-매우 높은 것으로 나타났고, 고전검사이론에 근거한 문항변별도 지수는 2개 블록 조건은 매우 낮음-낮음 사이였으며 3개 블록 조건부터 5개 블록 조건까지 중간에서 매우 높은 수준의 변별력을 지니는 것으로 산출되었다. 문항양호도 지수와 문항별 순위를 Table 8에 제시하였다.

매트릭스 역순

문항반응이론에 근거한 문항난이도 지수는 2개 블록이 점등되는 조건은 중간, 3개 블록 조건에서는 쉬움-중간, 4개와 5개 블록 조건에서는 모두 어려움으로 산출되었다. 고전검사이론에 근거한 난이도는 2개 블록 조건은 매우 쉬움-쉬움, 3개 블록 조건은 쉬움-중간으로 나타났지만, 4개 및 5개 블록 조건에서는 매우 어려움으로 나타났다. 중간-어려움, 5개 블록 조건에서는 어려움-매우 어려움으로 나타났다. 매트릭스 정순 과제와 유사하게 역순 과제에서도 4개 블록이 제시된 조건의 4번 문항의 난이도가 주변 문항보다 쉬운 것으로 나타나 문항이 수정될 필요가 있는 것으로 보인다.

문항변별도를 분석한 결과, 문항반응이론에 근거한 문항변별도 지수는 2개 블록조건은 중간-높은 수준으로 변별하였고, 3개 블록 조건부터 5개 블록 조건까지는 매우 높은 수준으로 나타났다. 고전검사이론에 근거한 문항변별도 지수는 2개 블록 조건은 낮음-중간 수준으로 나타났고, 3개, 4개 블록 조건에서는 높음, 그리고 마지막 5개 블록 조건은 중간 수준의 변별도 지수를 갖는 것으로 나타났다

Table 7. NWR item difficulty and item discrimination indices by IRT and CTT

NWR	Item number	IRT		CTT		Rank by item difficulty
		Difficulty	Discrimination	Difficulty	Discrimination	
2 syllables	1	-6.077	.400	.914	.106	19
	2	-9.459	.185	.850	.135	20
	3	-3.349	.636	.879	.192	18
	4	-3.026	.501	.808	.269	17
3 syllables	1	-1.159	.557	.649	.404	11
	2	-1.457	.640	.703	.317	13
	3	-2.839	.734	.869	.192	16
4 syllables	1	-1.939	1.019	.840	.279	14
	2	-.312	1.282	.578	.606	8
	3	-1.261	1.116	.760	.423	12
5 syllables	1	-.207	1.038	.546	.558	6
	2	-.232	1.686	.569	.635	7
	3	1.217	1.033	.262	.404	2
	4	-.603	1.222	.642	.529	10
6 syllables	1	.385	.960	.425	.500	5
	2	.601	1.399	.351	.548	4
	3	.697	1.260	.339	.538	3
	4	1.235	1.104	.249	.394	1

NWR=nonword repetition; IRT = item response theory; CTT=classical test theory.

Table 8. Matrix forward task item difficulty and item discrimination indices based on IRT and CTT

Matrix forward	Item number	IRT		CTT		Rank by item difficulty
		Difficulty	Discrimination	Difficulty	Discrimination	
2 blocks	1	-1.405	1.581	.832	.379	15
	2	-2.237	1.223	.900	.194	16
	3	-2.921	.863	.903	.194	17
	4	-3.068	.862	.913	.165	18
3 blocks	1	-.489	1.465	.628	.631	11
	2	-.815	1.806	.725	.583	14
	3	-.695	1.561	.683	.602	13
	4	-.627	1.697	.673	.660	12
4 blocks	1	.497	1.773	.362	.670	7
	2	.278	1.573	.430	.680	9
	3	.454	1.446	.385	.612	8
	4	-.269	1.273	.570	.641	10
	5	.919	1.872	.249	.612	4
5 blocks	1	.886	2.443	.239	.563	5
	2	1.046	2.917	.188	.534	3
	3	.810	2.356	.262	.621	6
	4	1.361	1.881	.155	.350	2
	5	1.387	2.878	.117	.330	1

IRT = item response theory; CTT = classical test theory.

Table 9. Matrix backward task item difficulty and item discrimination indices based on IRT and CTT

Matrix backward	Item number	IRT		CTT		Rank by item difficulty
		Difficulty	Discrimination	Difficulty	Discrimination	
2 blocks	1	-1.963	1.498	.898	.216	18
	2	-1.512	1.260	.818	.341	17
	3	-1.366	.997	.761	.386	16
	4	-1.208	1.327	.777	.420	15
3 blocks	1	-.676	1.480	.682	.568	13
	2	-.084	1.944	.542	.784	12
	3	-.835	1.691	.731	.545	14
	4	.162	1.488	.470	.705	10
4 blocks	1	.781	2.077	.284	.636	9
	2	.931	2.053	.246	.545	8
	3	1.024	2.151	.220	.511	7
	4	-.013	1.776	.519	.727	11
	5	1.165	2.200	.186	.443	6
5 blocks	1	1.322	2.707	.136	.398	4
	2	1.451	1.740	.155	.398	3
	3	1.251	2.538	.155	.375	5
	4	1.501	1.729	.148	.341	1
	5	1.468	2.265	.125	.330	2

IRT = item response theory; CTT = classical test theory.

다. 문항반응이론에 입각한 문항변별도 지수가 고전검사이론에 근거한 문항변별도 지수보다 문항변별도가 전반적으로 높게 나타났다. 문항양호도 지수와 문항 별 순위를 Table 9에 제시하였다.

문장 따라말하기

문장 따라말하기 과제의 문항난이도 지수를 문항반응이론에 근거하여 분석한 결과, 대부분의 문항이 매우 쉬움에서 중간 수준의 난이도로 추정되었다. 그러나 3어절 단문의 9번 문항난이도가 이상치로 나타나 추가 분석을 실시하여 내적일관성 지수를 살펴본 결과, 전체 문항에 대한 Cronbach's α 는 .87인 반면, 9번 문항 제거 시 신뢰도는 .88로 높아지는 것으로 나타나 문항을 제거하는 것이 본 검사의 내적일관성을 약간 높이는 것으로 보인다. 고전검사이론 모형에 근거한 문항난이도 지수도 앞선 결과와 유사하게 매우 쉬움부터 중간 수준까지 고르게 난이도가 추정되었다. 그러나 5어절의 문장들 간의 난이도가 단계 간 혼재되어 있는 것으로 나타나 5어절 문항을 전반적으로 재구성할 필요가 있는 것으로 나타났다. 특히 5어절 접속복문보다 5어절 단문이 순위가 높은 것으로 나타났다. 접속복문의 경우 두개의 문장의 시간적 순서나 원인과 결과, 조건과 결과로 구성되어 구문구조는 단문보다 복잡해도 내용을 논리적으로 이해하는데 어려움이 없는 문장들이다. 5어절 단문의 경우 주

어-목적어-서술어 구조에 관형어와 부사어 같은 목적어나 서술어를 꾸며주는 어휘가 추가된 형태로 문장을 이해하는데 있어서 부수적인 정보가 제시된 문장들이었다. 본 연구에서는 만 4-8세 아동의 경우 부수적인 정보를 포함한 단문보다 구문구조가 복잡하여도 관형어나 부사어 없이 논리적으로 내용이 전개되는 접속복문이 더 쉽게 느껴진 것으로 나타났다.

문항변별도 지수를 살펴본 결과, 문항반응이론에 근거하였을 때에는 모든 문항이 중간에서 높은 수준의 변별도를 지닌 반면, 고전 검사이론의 경우 3어절 단문과 5어절 접속복문의 변별도 지수는 매우 낮음에서 낮은 수준으로 해석되었고, 5어절 단문은 낮음-높음, 5어절 내포복문은 중간-높음 수준으로 나타나 전반적인 문항의 재배치가 필요한 것으로 보인다.

문장 따라말하기의 문항난이도와 문항변별도를 살펴보았을 때, 앞에서 소개된 다른 작업기억 과제들보다 문항난이도가 낮고 변별력도 낮은 것으로 해석된다. 문항양호도 지수와 문항 별 순위를 Table 10에 제시하였다.

2차 예비검사의 문항분석

1차 예비검사를 통해 문항양호도 지수가 비슷하여 중복된 문항들은 제거하고, 난이도를 기준으로 문항을 재배치하여 2차 예비검사를 위한 문항을 선정하였다. 이후, 2차 예비검사를 실시하였으며, 분석 결과, 전반적으로 2차 예비검사 자료는 고전검사이론에 근거한 추정치가 문항이 올라갈수록 문항난이도와 문항변별도 지수가 높아지는 형태로 나타나 문항반응이론 모형에 근거한 추정치에 비해 안정적으로 나타났다. 과제별 고전검사이론에 근거한 문항양호도 지수는 Figure 1-4에 제시하였고, 문항반응이론에 근거한 문항난이도 및 문항변별도 지수는 Appendix 5에 표로 제시하였다.

비단어 따라말하기 과제의 문항난이도 지수는 2-3음절은 쉬움, 4음절은 중간, 5음절은 중간-어려움, 6음절은 어려움으로 나타나 전반적으로 문항이 올라갈수록 난이도가 높아지는 형태로 추정되었다. 문항변별도의 경우, 2음절에서 4음절까지는 대체로 중간 수준으로 나타났으나 5음절 이상부터는 높은 수준으로 나타났다.

매트릭스 정순과 역순 과제의 경우 Figure 2와 Figure 3에 제시된 것과 같이 전반적으로 검사를 진행할 수록 문항난이도와 문항변별도 지수가 높아지는 현상이 뚜렷한 것으로 나타났다. 매트릭스 정순 과제의 경우 문항난이도 지수는 2개 블록 조건은 쉬움에서 중간 정도의 난이도로 추정되었고, 3개 블록 조건은 중간과 어려움 수준으로 추정되었다. 4-5개 블록 조건은 어려움과 매우 어려움 수준의 난이도로 추정되어 문항이 높아질수록 난이도도 함께 높아지는 것으로 나타났다. 그리고 문항변별도 지수를 언어적으로 표현

Table 10. SR task item difficulty and discrimination indices based on IRT and CTT

SR	Item number	IRT		CTT		Rank by item difficulty
		Difficulty	Discrimination	Difficulty	Discrimination	
3 words SS	1	-3.429	1.575	.985	.043	29
	2	-6.721	.515	.966	.057	35
	3	-4.055	.670	.927	.115	32
	4	-3.456	1.130	.966	.072	30
	5	-3.233	2.871	.995	.014	28
	6	-3.835	1.041	.971	.057	31
	7	-4.608	.785	.966	.057	33
	8	-5.728	.727	.980	.043	34
	9	32.283	-.109	.971	0	1
5 words SS	1	-.821	2.088	.754	.521	13
	2	-1.272	1.287	.788	.333	20
	3	-.823	1.646	.730	.550	14
	4	-1.235	1.446	.798	.362	19
	5	-0.366	1.218	.596	.536	3
	6	-.755	1.169	.677	.449	9
	7	-.731	2.053	.730	.550	8
	8	-.700	1.190	.668	.492	6
	9	-1.554	1.279	.831	.289	25
5 words CS	1	-.82	1.583	.725	.507	12
	2	-1.357	1.334	.807	.347	22
	3	-1.353	1.341	.807	.376	21
	4	-1.540	1.108	.807	.362	24
	5	-2.333	1.188	.908	.188	27
	6	-1.010	1.613	.769	.478	17
	7	-1.453	.944	.769	.347	23
	8	-.957	1.777	.769	.536	15
	9	-2.199	.891	.850	.260	26
5 words ES	1	-.283	1.222	.576	.608	2
	2	-.546	1.184	.634	.521	5
	3	-.764	1.151	.677	.478	10
	4	-1.042	1.530	.769	.463	18
	5	-.498	1.518	.644	.623	4
	6	-.803	1.324	.701	.492	11
	7	-.962	1.496	.750	.492	16
	8	-.711	1.607	.701	.507	7

SS=simple sentence; CS=conjoined sentence; ES=embedded sentence; IRT=item response theory; CTT=classical test theory.

하였을 때 대부분의 문항이 매우 높음으로 해석되었다. 그러나 블록 4개 점등되는 조건의 5번부터 5개 블록 조건의 문항은 변별도가 .2 이하로 매우 낮아서 해당 문항을 최종 문항에 포함시키기엔 신뢰롭지 못한 것으로 보인다.

그리고 매트릭스 역순 과제의 문항난이도 지수는 2개 블록 조건은 쉬움에서 중간, 3개 블록 조건은 중간에서 어려움, 4개 블록 조

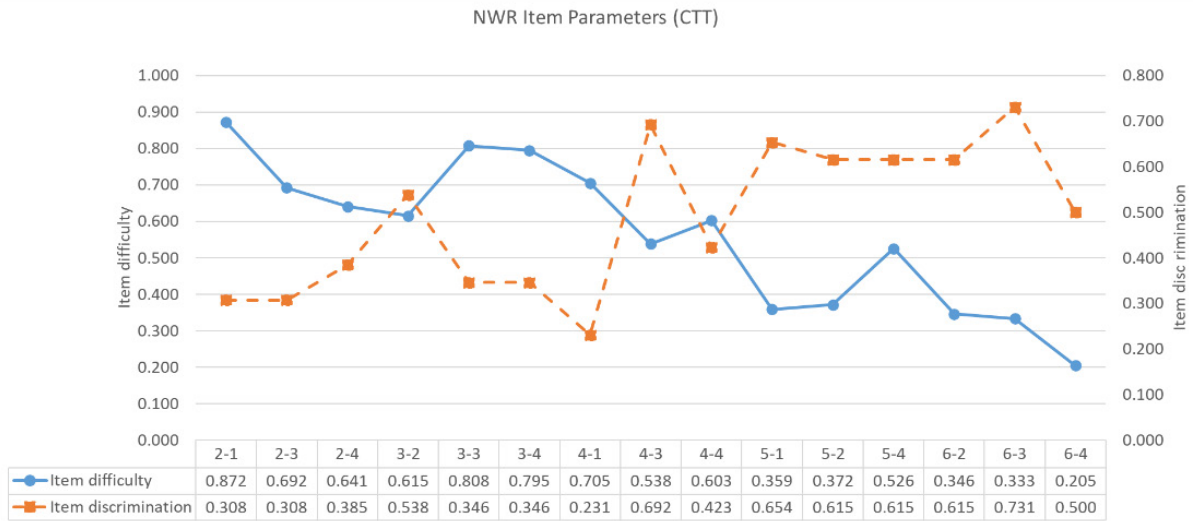


Figure 1. Item parameters of nonword repetition task.
CTT = classical test theory.

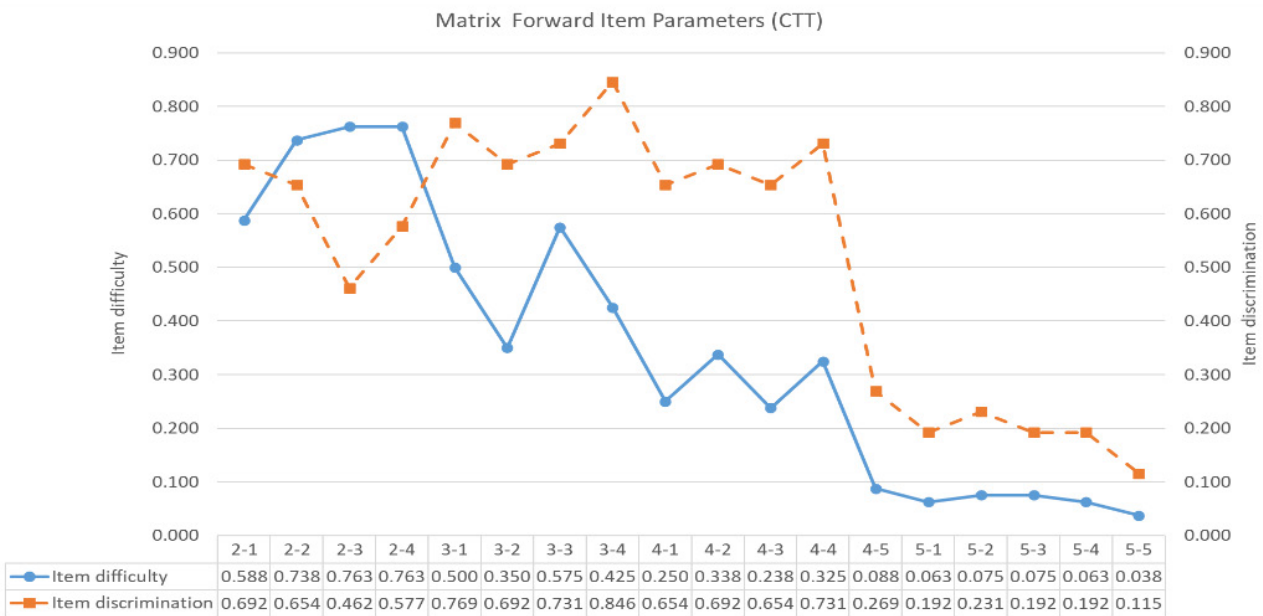


Figure 2. Item parameters of Matrix forward task.
CTT = classical test theory.

건부터는 매우 어려운 수준으로 해석되었다. 문항변별도 지수를 살펴보면, 첫 문항부터 4개 블록 조건의 1번 문항까지는 매우 높게 추정되었으나 그 이후 문항부터는 문항변별도 지수가 낮게 추정되어 변별도가 없는 문항으로 나타났다.

문장 따라말하기 과제의 경우 전체 문항의 문항난이도 지수가 비슷하게 나타났다(Figure 4). 문항난이도 지수를 언어적으로 해석하였을 때 문항 3어절 단문은 매우 쉬움, 5어절 단문은 매우 쉬움에

서 쉬움, 5어절 접속복문은 쉬움, 5어절 내포복문은 쉬움에서 중간 수준이었다. 문항변별도 지수를 살펴보면, 3어절 단문은 변별도가 매우 낮게 나타났으나, 5어절 단문, 접속복문, 내포복문은 모두 변별도가 매우 높게 추정되었다.

공인타당도

작업기억 측정과제의 공인타당도를 확보하기 위하여 수용·표현

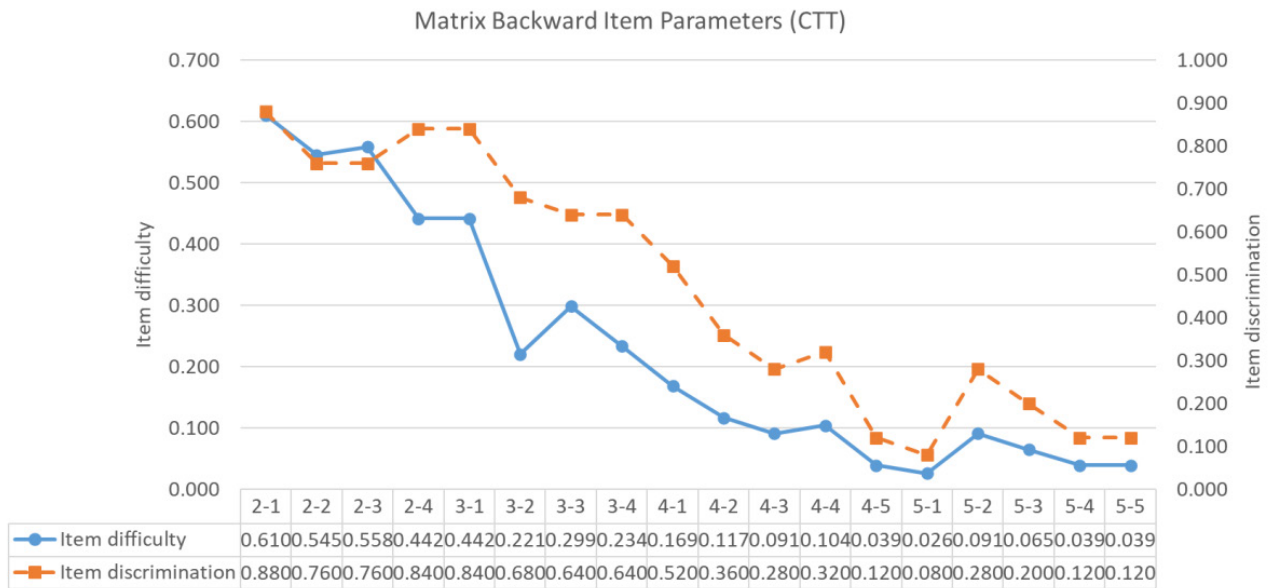


Figure 3. Item parameters of Matrix backward task. CTT=classical test theory.

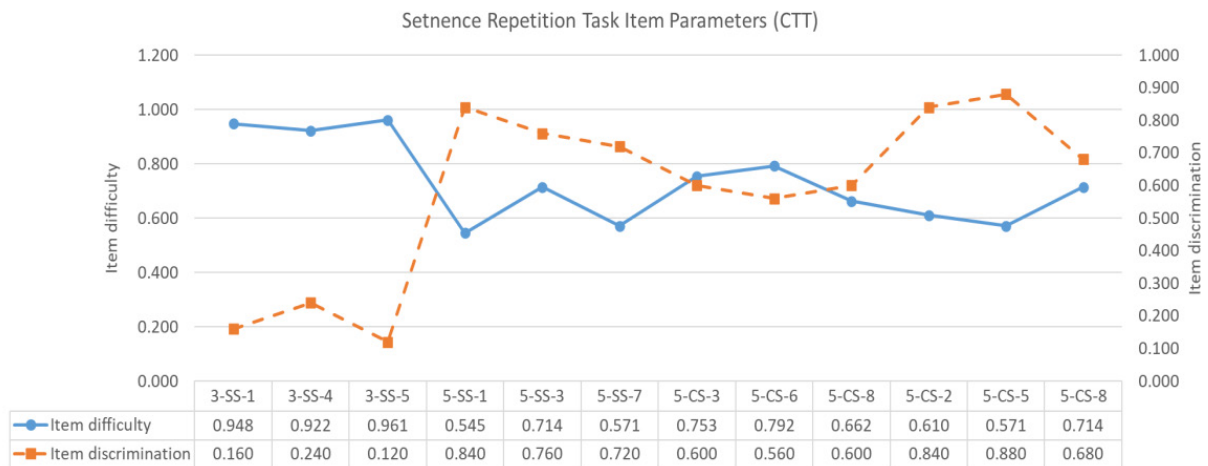


Figure 4. Item parameters of sentence repetition task. SS= simple sentence; CS= conjoined sentence; ES= embedded sentence. CTT=classical test theory.

어휘 검사(REVT; Kim et al., 2009)와의 상관분석을 실시하였다. 2차 예비연구에 참여한 아동을 대상으로 수용어휘 원점수와 상관을 분석한 결과, 비단어 따라말하기 과제는 .529 ($N=78, p=.000$), 매트릭스 정순 과제는 .569 ($N=80, p=.000$) 매트릭스 역순 과제는 .545 ($N=77, p=.000$), 문장 따라말하기 과제는 .729 ($N=77, p=.000$)로 모두 유의한 정적 상관이 나타났다. 그리고 표현어휘 원점수는 비단어 따라말하기 과제와 .594 ($N=78, p=.000$), 매트릭스 정순 과제와 .459 ($N=80, p=.000$) 매트릭스 역순 과제와 .534 ($N=77, p=.000$), 문장 따라말하기 과제와 .691 ($N=77, p=.000$)의 유의한

정적 상관을 보여 작업기억 측정과제의 공인타당도를 확보하였다.

논의 및 결론

아동의 언어능력을 다면적으로 평가하기 위해서는 전통적인 언어평가법에 작업기억 평가와 같은 언어발달에 영향을 미치는 능력을 함께 평가하는 것이 중요하다. 그러나 여러 연구에서 통일되지 않은 다양한 검사도구를 사용하여 아동의 작업기억을 평가하고 있어서 임상 현장에서의 활용을 위해서는 우선 과제의 문항양호도

검증을 실시하여 아동에게 적합한 문항으로 구성된 작업기억 과제를 제시할 필요가 있었다. 이에 본 연구에서는 연구와 임상 현장 간의 거리를 메우기 위한 첫 번째 단계로, 기 발표된 연구를 통해 신뢰도를 확보한 작업기억 검사도구 4개를 선정하여, 문항반응이론 모형 및 고전검사이론을 기반으로 문항양호도를 분석하였다. 이를 통해 작업기억 측정과제의 연령 별, 집단 별 수행치 및 문항난이도, 문항변별도를 제시하여 측정도구를 타당화하였다. 이에 대한 논의점은 다음과 같다.

문항난이도를 살펴본 결과, 언어정보와 관련된 작업기억을 측정하는 비단어 따라말하기와 문장 따라말하기 과제는 대체적으로 쉬운 과제로 나타났다. 과제 내 단계 별 문항난이도는 문항 번호가 올라갈수록 문항난이도도 올라가도록 배치되었으나 쉬운 문항과 어려운 문항이 순차적으로 분포되어 있지는 않았다. 기존 문항은 5어절 단문, 5어절 접속복문, 5어절 내포복문 순으로 문항이 제시되었는데 문항난이도와 문항변별도 지수를 분석한 결과 5어절 단문의 일부 문항난이도가 5어절 접속문보다 문항난이도가 높았고, 제일 문항난이도가 높다고 가정하여 앞쪽에 배치한 내포문과도 문항난이도가 유사하게 나타났다. 이는 5어절 단문 “아빠가 동생한테 맛있는 빵을 줘요.”의 경우 어순을 바르게 산출해야 점수를 획득할 수 있는데 어순을 다르게 하여도 문장 의미에 영향을 미치지 않는 점이 반영된 결과로 보인다. 반면 5어절 내포문 “아빠가 청소를 하는 엄마를 불러요.”와 5어절 접속문 “아빠가 힘들어서 하품을 크게 해요.”는 단문과 달리 논리적인 구조를 포함하고 있어서 어순을 바꾸어 산출할 확률이 낮다.

1차 예비검사를 통해 산출된 문항난이도를 반영하여 문항을 재배치하고 불필요한 문항을 삭제하여 2차 검사를 실시한 결과, 문항 통과율과 문항난이도 및 변별도가 적절하게 재 분포된 것을 확인할 수 있었다. 그러나 더 많은 표본을 포함하기 위해 전국 단위의 표준화 연구를 진행할 경우 추가로 수정되어야 할 문항이 존재할 것으로 보여 보충 문항이 필요하다고 여겨진다.

비언어적 작업기억을 측정하기 위해 실시한 매트릭스 정순, 역순 과제 역시 첫 번째 예비검사 결과 난이도가 쉬운 문항부터 어려운 문항까지 모두 분포하고 있는 것으로 나타났다. 그리고 전반적인 두 매트릭스 과제의 수행은 비단어 따라말하기와 문장 따라말하기 과제보다는 낮아서 난이도 측면에서 비언어정보 과제가 더 어려운 것으로 나타났다. 그러나 앞에 문항이 바로 이어서 나오는 문항과 자극 제시 순서 및 위치가 유사하여 점화효과(priming effect)가 나타날 수 있는 문항이 있었다. 그리고 자극이 제시되는 양이 많아져도, 제시된 자극들이 자칫 패턴을 형성할 경우 아동의 수행력이 높아져 문항의 난이도에 영향을 주는 것으로 보였다. 이에 자극 제시 순

서와 위치를 수정하여 2차 예비검사를 실시한 결과 문항의 난이도와 변별도가 문항이 높아질수록 극명하게 높아졌으며, 문항 통과율 역시 난이도가 올라가며 뚜렷하게 낮아지는 것으로 나타났다. 1차, 2차 예비검사를 통해 보다 안정적인 결과를 도출할 수 있는 문항으로 구성된 매트릭스 과제가 완성되어 임상 현장에서 언어능력 평가만으로 아동의 발달을 평가하기 어려울 때 본 매트릭스 검사를 사용하여 아동의 정보처리 능력을 고려한 평가를 실시하는 것이 유용할 것이다.

본 연구는 오래전부터 지속적으로 논의되고 있는 작업기억 평가를 임상 현장에 적용할 수 있도록 발판을 마련했다는 점에서 의의가 있다. 연구와 임상 현장 간의 간격을 메우고자 아동 평가 시 참고할 수 있는 언어적, 비언어적 작업기억 측정과제의 연령 별, 성별, 집단 별 수행에 대한 기초통계자료를 제시하였고 문항 수를 대폭 줄임으로써 아동 특성을 고려한 작업기억 측정과제를 제공하였다. 이는 이중언어 아동이나 다문화 아동 또는 언어를 제한적으로 사용하는 다양한 임상군을 대상으로 그들의 정보처리적 특성을 파악하는데 유용하게 활용될 것이다.

본 연구의 제한점과 후속연구를 위한 제언은 다음과 같다. 다수의 연구에서 신뢰도를 보고한 네 개의 작업기억 측정과제의 문항 분석을 실시하여 신뢰로운 결과를 도출할 수 있는 도구를 개발하고자 두 번의 예비검사를 통해 문항을 수정하고 재배치하였으나, 타당도 연구 특성상 연령 별, 집단 별 표본의 크기가 충분히 크지 못했다. 본 연구에서 사용한 문항반응이론으로 추정된 문항난이도와 변별도의 해석에 관해 논의하자면, 문항반응이론의 장점에도 불구하고, 가장 큰 장애물은 표본 크기에 대한 까다로운 요건이다. 표본 크기에 대한 Lord (1968)의 첫 시뮬레이션 연구와 후속연구 (Patsula & Gessaroli, 1995; Tang, Way, & Carey, 1993; Yoes, 1995)에서는 최소 표본 크기를 1,000으로 제안하였다. 한편, Sahin과 Anil (2017)은 대학교 1학년 6,288명을 대상으로 50개 문항으로 구성된 실제 영어 어학검사 응답자료를 이용하여, 3가지 문항반응이론 모형(1모수, 2모수, 3모수 모형)의 문항-모수 추정에 대한 표본 크기와 문항길이의 영향을 조사한 시뮬레이션 연구를 시도하였다. 연구 결과, 일차원성이 가정되고 주변최대우도추정(Marginal Maximum likelihood estimation)을 이용하는 경우, 문항난이도 모수를 정확하게 추정하기 위해 2모수 모형에 적절한 표본 크기는 문항 수가 10개는 750, 20개는 500, 30개는 250으로 제안하였다. 실제 자료를 이용한 표본 크기에 대한 선행연구 결과와 비교해 볼 때, 2차 예비검사의 표본 크기는 작기 때문에 해석상 주의가 요구되며, 추후 표본을 추가하여 타당도를 보완할 필요가 있다.

이러한 제한점에도 불구하고 본 연구결과는 문항반응이론이 임

상 장면에서 활용할 수 있는 도구를 제작하는데 유용한 정보를 제공할 수 있음을 보였다는데 의의가 있다. 연령 별 표본 크기가 8에서 30으로 매우 작지만, 현 단계에서는 연령 별 반응정도를 탐색하는 것이 주된 목적이었기 때문에 과제 별 최종 문항(본 문항) 선정 시 이를 고려해야 할 것이다. 문항반응이론은 표본의 특성에 따라 문항이 가진 자체 특성이 변하지 않는다는 점에서 교육평가 및 보건 분야에서 널리 사용되고 있는 검사이론이다. 언어병리학 분야에서도 관심을 갖고 문항반응이론에 근거한 평가문항이 개발될 수 있도록 지속적인 연구가 필요하다. 현실점에서는 문항반응이론의 기본 가정과 표본 크기의 특성을 만족하는 검사도구 제작이 어려울 수 있지만, 고전검사이론에 근거한 자료를 지속적으로 축적하여 궁극적으로는 문항반응이론 모형에 근거한 문항으로 구성된 작업 기억 평가도구를 개발해야 할 것이다.

후속연구에서는 현재 가장 많이 표집한 만 4-6세 아동으로 타겟 연령을 선정하여 본 검사 자료를 효율적으로 수집하여야 하며, 조음음운장애, 읽기장애 등을 포함하여 언어장애를 판별할 수 있는 진단 기준점을 산출하고 진단 도구로써 유용성을 검증할 것을 제안한다.

REFERENCES

- Adams, A. M., & Gathercole, S. E. (1995). Phonological working memory and speech production in preschool children. *Journal of Speech, Language, and Hearing Research, 38*(2), 403-414.
- Alloway, T. P., Rajendran, G., & Archibald, L. M. (2009). Working memory in children with developmental disorders. *Journal of Learning Disabilities, 42*(4), 372-382.
- Archibald, L. M., & Gathercole, S. E. (2006). Nonword repetition: a comparison of tests. *Journal of Speech, Language, and Hearing Research, 49*(5), 970-983.
- Baddeley, A. D. (1997). *Human memory: theory and practice*. Psychology Press.
- Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158-173.
- Baker, F. B. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation.
- Berninger, V. W., Nielsen, K. H., Abbott, R. D., Wijsman, E., & Raskind, W. (2008). Writing problems in developmental dyslexia: under-recognized and under-treated. *Journal of School Psychology, 46*(1), 1-21.
- Boudreau, D. (2008). Narrative abilities: advances in research and implications for clinical practice. *Topics in Language Disorders, 28*(2), 99-114.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Memory and Language, 19*(4), 450-466.
- Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41*(5), 1136-1146.
- Ebel, R. L. (1965). Confidence weighting and test reliability. *Journal of Educational Measurement, 2*(1), 49-57.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Gathercole, S. E. (2006). Nonword repetition and word learning: the nature of the relationship. *Applied Psycholinguistics, 27*(4), 513-543.
- Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: a longitudinal study. *Journal of Memory and Language, 28*(2), 200-213.
- Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: is there a causal connection?. *Journal of Memory and Language, 29*(3), 336-360.
- Gathercole, S. E., & Baddeley, A. D. (1993). Phonological working memory: a critical building block for reading development and vocabulary acquisition?. *European Journal of Psychology of Education, 8*(3), 259-272.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: principles and application*. Norwell, MA: Kluwer Academic Publishers.
- Han, W. J., & Yim, D. (2016). The relationship among complex sentence production, implicit learning, and working memory based on structural priming method in children with specific language impairment. *Communication Sciences & Disorders, 21*(4), 590-604.
- Hwang, M. A., Lim, J. A., Choi, K. S., Ko, S. H., Choi, S. Y., Kim, J. H., Jeon, M. Y. (2016). Inhibition of interference during word reading in children with attention deficit hyperactive disorder. *Communication Sciences & Disorders, 21*(1), 121-130.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review, 99*(1), 122-149.
- Kim, S. Y., & Yim, D. (2015). Study of working memory intervention in children with delay in vocabulary development: effects on working memory and language ability. *Communication Sciences & Disorders, 20*(4), 469-489.
- Kim, Y. T. (1997). Study on utterance length in 2-4 year-old Korean children. *Korean Journal of Communication & Disorders, 2*(1), 5-25.
- Kim, Y. T., Hong, G. H., Kim, K. H., Jang, H. S., & Lee, J. Y. (2009). *Receptive*

- & expressive vocabulary test (REVT), Seoul: Seoul Community Rehabilitation Center.
- Kim, Y. T., Sung, T. J., & Lee, Y. K. (2003). *Preschool receptive-expressive language scale (PRES)*. Seoul: Seoul Community Rehabilitation Center.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *Journal of Memory and Language*, 30(5), 580-602.
- Lawley, D. N. (1943). XXIII. On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 273-287.
- Lee, Y. S., Kwon, K. A., & Kim, S. I. (1979). *Sentence development of Korean children (I)*. Seoul: Korean Education Development Institute.
- Lee, S. K., Jung, S. H., Lee, E. H., & Lee, J. H. (1972). An investigation of vocabulary in young children. *Journal of Educational Studies*, 1, 43-133.
- Leonard, L. B., Elis Weismer, S., Miller, C. A., Francis, D. J., Tomblin, J. B., & Kail, R. V. (2007). Speed of processing, working memory, and language impairment in children. *Journal of Speech, Language, and Hearing Research*, 50(2), 408-428.
- Lord, F. M. (1968). An analysis of the Verbal Scholastic Aptitude Test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement*, 28(4), 989-1020.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109(1), 35-54.
- Marton, K., & Schwartz, R. G. (2003). Working memory capacity and language processes in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 36(5), 1138-1153.
- Monaco, M., Costa, A., Caltagirone, C., & Carlesimo, G. A. (2013). Forward and backward span for verbal and visuo-spatial data: standardization and normative data from an Italian adult population. *Neurological Sciences*, 34(5), 749-754.
- Montgomery, J. W. (1995). Examination of phonological working memory in specifically language-impaired children. *Applied Psycholinguistics*, 16(4), 355-378.
- Moon, S. B., & Byun, C. J. (2003). *Korean Kaufman assessment battery for children (K-ABC)*. Seoul: Hakjisa.
- Munson, B., Edwards, J., & Beckman, M. E. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research*, 48(1), 61-78.
- Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, 48(5), 1033-1047.
- Muse, A. E. (2003). *Do individual differences in lexical representations or speech output account for relations between nonword repetition or vocabulary?* (Master's thesis). Florida State University, Tallahassee, USA.
- Oh, D. Y., & Yim, D. (2013). Non-word repetition and sentence repetition performance in 2-3 years old late talkers and normal children. *Communication Sciences & Disorders*, 18(3), 277-287.
- Partchev, I., Partchev, M. I., & Suggests, M. A. S. S. (2017). Package 'irtoys'. A collection of functions related to item response theory (IRT). R package version.3.5.3.
- Patsula, L., & Gessaroli, M.E. (1995). A comparison of item parameter estimates and ICCs produced with TESTGRAF and BILOG under different test lengths and sample sizes. *Applied psychological measurement*, 13, 57-75.
- Şahin, A., & Anil, D. (2017). The effects of test length and sample size on item parameters in item response theory. *Educational Sciences: Theory & Practice*, 17, 321-335.
- Schumacker, R. (2019) Psychometric packages in R. *Measurement: Interdisciplinary Research and Perspectives*, 17(2), 106-112.
- Sung, T. J. (1991). *Introduction to Item Response Theory*. Seoul: Yangseowon.
- Sung, T. J. (2002). *Modern educational evaluation*. Seoul: Hakjisa
- Sung, T. J. (2016). *Item response theory model application*. Seoul: Kyowookkwahaksa.
- Swanson, H. L., & Berninger, V. (1995). The role of working memory in skilled and less skilled readers' comprehension. *Intelligence*, 21(1), 83-108.
- Tang, K. L., Way, W. D., & Carey, P. A. (1993). The effect of small calibration sample sizes on TOEFL IRT-based equating. *ETS Research Report Series*, 1993(2), i-38.
- Team, R. D. C. (2013). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing; Vienna, Austria.
- Wilson, J. L., Scott, J. H., & Power, K. G. (1987). Developmental differences in the span of visual memory for pattern. *British Journal of Developmental Psychology*, 5(3), 249-255.
- Yang, Y. H., & Yim, D. (2018). The role of executive function for vocabulary acquisition and word learning in preschool-age children with and without vocabulary delay. *Communication Sciences & Disorders*, 23(1), 1-17.
- Yang, Y. H., Yim, D., & Bae, K. R. (2015). Predictors of word learning in chil-

- dren with specific language impairment. *Communication Sciences & Disorders*, 20(1), 1-12.
- Yang, Y. H., Yim, D., Kim, S. Y., & Han, J. Y. (2013). The relationship among receptive vocabulary, non-word repetition, and quick incidental learning in preschoolers with and without delay in vocabulary development. *Communication Sciences & Disorders*, 18(4), 379-391.
- Yim, D., Kim, Y. T., & Yang, Y. H. (2016). Exploring the utility of verbal and visuospatial working memory for identifying children with language impairment. *Communication Sciences & Disorders*, 21(2), 193-205.
- Yim, D., Yang, Y. H., & Kim, S. Y. (2015). Domain-specific working memory performance in children with and without specific language impairment. *Communication Sciences & Disorders*, 20(1), 13-23.
- Yoes, M. (1995). *An updated comparison of micro-computer based item parameter estimation procedures used with the 3-parameter IRT model*. Saint Paul, MN: Assessment Systems Corporation..
- Yoon, H. J. (2015). Prediction of reading comprehension in early and late elementary grades: contribution of word decoding, vocabulary and syntactic knowledge. *Communication Sciences & Disorders*, 20(4), 536-546.

Appendix 1. Number of participants by age in each group at 1st and 2nd preliminary studies

1st Preliminary study					2nd Preliminary study				
Group	Age	Sex	N	%	Group	Age	Sex	N	%
NL (N=248)	4	M	26	10.4	NL (N=56)	4	M	10	17.9
		F	24	9.6			F	12	21.4
	5	M	49	19.7		5	M	9	16.0
		F	59	23.7			F	14	25.0
	6	M	40	16.1		6	M	5	8.9
		F	42	16.9			F	6	10.8
	7	M	1	.6		7	M	0	0
		F	2	1			F	0	0
	8	M	2	1		8	M	0	0
		F	3	1			F	0	0
LI (N=81)	4	M	10	12.3	LI (N=28)	4	M	7	21.9
		F	5	6.2			F	4	12.5
	5	M	16	19.8		5	M	9	28.1
		F	11	13.6			F	3	9.4
	6	M	16	19.8		6	M	5	15.6
		F	11	13.6			F	4	12.5
	7	M	4	4.9		7	M	0	0
		F	4	4.9			F	0	0
	8	M	1	1.2		8	M	0	0
		F	3	3.7			F	0	0

NL=children with normal language development; LI=children with language impairment.

Appendix 2. Comparison of original items vs. revised items on NWR

NWR	Item number	Item	Modification	Revised version
2 syllable	1	누베		누베
	2	마뚜	Deleted	
	3	조나		조나
	4	퍼틱		퍼틱
3 syllable	1	버줍디	Deleted	
	2	까다굳		까다굳
	3	모단기		모단기
	4	니아토		니아토
4 syllable	1	투보가인		투보가인
	2	머구냥뿔	Deleted	
	3	푸가태지		푸가태지
	4	드반거노		드반거노
5 syllable	1	조매누버리		조매누버리
	2	레빌애티머		레빌애티머
	3	누빈재구밈	Deleted	
	4	바줍다거니		바줍다거니
6 syllable	1	무지다바리노	Deleted	
	2	로밈띠르저니		로밈띠르저니
	3	보마데낭까두		보마데낭까두
	4	미기돌아캐바		미기돌아캐바

NWR=nonword repetition.

Appendix 3. Comparison of original items vs. revised items on Matrix forward and backward

Condition	Item number	Modification	
		Matrix forward	Matrix backward
2 blocks	1		
	2	Order and location modified	Order and location modified
	3	Order and location modified	Order and location modified
3 blocks	4		
	1	Order and location modified	
	2	Order and location modified	Order and location modified
	3		Order and location modified
4 blocks	4		
	1		
	2		
	3	Order and location modified	
	4	Order and location modified	Order and location modified
5 blocks	5		
	1		Order and location modified
	2		
	3	Order and location modified	
	4		Order and location modified

Appendix 4. Comparison of original items vs. revised items on SR

Sentence repetition	Item number	Item	Modification	Revised version
3 words SS	1	그림을 연필로 그려요.		그림을 연필로 그려요.
	2	색종이를 풀로 붙여요.	Deleted	
	3	머리를 삼푸로 감아요.	Deleted	
	4	동생이 그림책을 봐요.		동생이 그림책을 봐요.
	5	엄마가 치마를 입어요.		엄마가 치마를 입어요.
	6	아빠가 신문을 읽어요.	Deleted	
	7	아가가 집에서 놀아요	Deleted	
	8	친구가 차에서 내려요.	Deleted	
	9	엄마가 시장에서 와요.	Deleted	
5 words SS	1	동생이 작은 칫솔로 이빨을 닦아요.	Order changed	엄마가 청소를 하고 걸레를 빨아요 (originally 5-word CS item 3)
	2	엄마가 커다란 비누로 빨래를 해요.	Deleted	
	3	친구가 작은 가위로 종이를 잘라요.	Deleted	
	4	아빠가 동생한테 맛있는 빵을 줘요.	Order changed	아빠가 힘들어서 하품을 크게 해요. (originally 5-word CS item 6)
	5	아빠가 아가한테 작은 공을 던져요.	Deleted	
	6	아빠가 아가한테 예쁜 옷을 사줘요.	Deleted	
	7	친구가 밖에서 그네를 재밌게 타요.	Order changed	동생이 추우면 엄마가 이불을 줘요. (originally 5-word CS item 8)
	8	엄마가 부엌에서 밥을 맛있게 해요.	Deleted	
	9	엄마가 시장에서 사과를 조금 사요.	Deleted	
5 words CS	1	친구가 목욕을 하고 머리를 감아요.	Deleted	
	2	동생이 바지를 입고 양말을 신어요.	Deleted	
	3	엄마가 청소를 하고 걸레를 빨아요.	Order changed	동생이 작은 칫솔로 이빨을 닦아요. (originally 5-word SS item 1)
	4	아가가 아파서 주사를 많이 맞아요.	Deleted	
	5	친구가 배고파서 밥을 빨리 먹어요.	Deleted	
	6	아빠가 힘들어서 하품을 크게 해요.	Order changed	아빠가 동생한테 맛있는 빵을 줘요 (originally 5-word SS item 4)
	7	아가가 안자면 엄마가 우유를 줘요.	Deleted	
	8	동생이 추우면 엄마가 이불을 줘요.	Order changed	친구가 밖에서 그네를 재밌게 타요. (originally 5-word SS item 7)
	9	동생이 아프면 아빠가 약을 사와요.	Deleted	
5 words ES	1	친구는 수영을 하는 것을 좋아해요.	Deleted	
	2	친구는 노래를 하는 것을 싫어해요.	Revised and order changed	아빠가 청소를 하는 엄마를 불러요 (originally 5-word ES item 5)
	3	아가는 그림을 그리는 것을 못해요.	Deleted	
	4	친구가 잠을 자는 강아지를 깨워요.	Deleted	
	5	아빠가 청소를 하는 엄마를 불러요.	Order changed	동생은 주사를 맞는 것을 싫어해요 (originally 5-word ES item 2)
	6	엄마가 머리를 빨리 감으라고 해요.	Deleted	
	7	아빠가 방을 깨끗이 닦으라고 해요.	Deleted	
	8	아빠가 창문을 세게 닫으라고 해요.		아빠가 창문을 세게 닫으라고 해요.

SR=sentence repetition; SS=simple sentence; CS=conjoined sentence; ES=embedded sentence.

Appendix 5. Item parameters for revised NWR, Matrix forward, Matrix backward, and SR tasks based on IRT and CTT

Working memory	Condition	Difficulty	Discrimination	Difficulty	Discrimination
NWR	Syllable-2 1	-1.914	1.277	0.872	0.308
	Syllable-2 3	-1.356	0.653	0.692	0.308
	Syllable-2 4	-1.225	0.500	0.641	0.385
	Syllable-3 2	-0.544	1.051	0.615	0.538
	Syllable-3 3	-2.009	0.805	0.808	0.346
	Syllable-3 4	-1.712	0.922	0.795	0.346
	Syllable-4 1	-1.828	0.504	0.705	0.231
	Syllable-4 3	-0.135	1.521	0.538	0.692
	Syllable-4 4	-0.726	0.622	0.603	0.423
	Syllable-5 1	0.530	1.653	0.359	0.654
	Syllable-5 2	0.535	1.335	0.372	0.615
	Syllable-5 4	-0.105	1.176	0.526	0.615
	Syllable-6 2	0.630	1.397	0.346	0.615
	Syllable-6 3	0.521	2.899	0.333	0.731
Syllable-6 4	1.105	1.967	0.205	0.500	
Matrix forward	Block-2 1	-0.244	1.476	0.588	0.692
	Block-2 2	-0.596	3.432	0.738	0.654
	Block-2 3	-0.902	1.805	0.763	0.462
	Block-2 4	-0.835	2.100	0.763	0.577
	Block-3 1	0.109	2.743	0.500	0.769
	Block-3 2	0.568	2.182	0.350	0.692
	Block-3 3	-0.117	2.484	0.575	0.731
	Block-3 4	0.338	1.989	0.425	0.846
	Block-4 1	1.003	1.764	0.250	0.654
	Block-4 2	0.679	1.640	0.338	0.692
	Block-4 3	0.840	3.556	0.238	0.654
	Block-4 4	0.615	2.629	0.325	0.731
	Block-4 5	1.444	31.723	0.088	0.269
	Block-5 1	1.709	3.922	0.063	0.192
	Block-5 2	1.743	2.711	0.075	0.231
	Block-5 3	1.744	2.705	0.075	0.192
	Block-5 4	1.811	2.936	0.063	0.192
	Block-5 5	2.253	2.310	0.038	0.115
Matrix backward	Block-2 1	-0.334	2.311	0.610	0.880
	Block-2 2	-0.133	1.850	0.545	0.760
	Block-2 3	-0.208	1.521	0.558	0.760
	Block-2 4	0.231	2.182	0.442	0.840
	Block-3 1	0.231	2.412	0.442	0.840
	Block-3 2	0.927	2.910	0.221	0.680
	Block-3 3	0.678	2.669	0.299	0.640
	Block-3 4	0.851	3.580	0.234	0.640
	Block-4 1	1.043	4.475	0.169	0.520
	Block-4 2	1.073	26.984	0.117	0.360
	Block-4 3	1.393	3.915	0.091	0.280
	Block-4 4	1.080	30.728	0.104	0.320
	Block-4 5	1.795	3.943	0.039	0.120
	Block-5 1	2.125	3.028	0.026	0.080
	Block-5 2	1.502	2.942	0.091	0.280
	Block-5 3	1.664	3.075	0.065	0.200
	Block-5 4	1.800	3.896	0.039	0.120
	Block-5 5	1.726	43.764	0.039	0.120
Sentence repetition	3_SS_1	-2.169	2.130	0.948	0.160
	3_SS_2	-1.737	2.723	0.922	0.240
	3_SS_3	-1.794	20.923	0.961	0.120
	5_SS_1	-0.115	2.067	0.545	0.840
	5_SS_2	-0.550	4.918	0.714	0.760
	5_SS_3	-0.208	1.948	0.571	0.720
	5_CS_1	-0.831	2.348	0.753	0.600
	5_CS_2	-0.929	2.989	0.792	0.560
	5_CS_3	-0.560	1.744	0.662	0.600
	5_ES_1	-0.320	2.240	0.610	0.840
	5_ES_2	-0.178	2.616	0.571	0.880
	5_ES_3	-0.819	1.558	0.714	0.680

NWR=nonword repetition; IRT = item response theory; CTT = classical test theory; SS= simple sentence; CS=conjoined sentence; ES=embedded sentence.

국문초록

아동용 작업기억 측정도구(K-WMM-C) 개발을 위한 문항분석

한지윤¹ · 임동선¹ · 석혜은²

¹이화여자대학교 언어병리학과, ²이화여자대학교 심리학과

배경 및 목적: 본 연구의 목적은 널리 사용되는 네 가지 작업기억 측정과제의 문항을 분석하고 작업기억 측정과제의 공인타당도를 확보하고자 하는 데 있다. **방법:** 본 논문에는 두 가지 예비 연구가 포함되어 있다. 첫 번째 예비연구는 만 4세에서 8세 사이의 아동 329명 (NL 248, LI 81)의 자료를 바탕으로 하였다. 비단어 따라말하기, 매트릭스 정순, 매트릭스 역순, 문장 따라말하기 과제가 사용되었다. 문항난이도는 문항반응이론과 고전검사이론을 바탕으로 분석하였다. 2차 예비연구에서는 만 4세에서 6세 사이의 84명의 아동을 대상으로 수정된 작업기억 과제를 실시한 자료를 사용하였다. 문항난이도 및 변별도를 분석하여 문항 양호도를 분석하였다. 마지막으로 표준화된 어휘검사 점수와 네 가지 작업기억 측정 간의 상관분석을 실시하여 공인 타당도의 증거를 제시하고자 하였다. **결과:** 대부분의 문항은 난이도 순으로 적절히 배치되었고, 난이도가 부적절한 일부 문항은 수정 및 제거되었다. 네 가지 작업기억 과제의 시작문항은 대부분의 아동이 90%이상 통과하는 것으로 나타났다. 두 번째 연구에서는, 문항난이도와 변별도를 산출하였으나, 1차 연구를 통해 문항을 수정했음에도 여전히 난이도와 변별도 지수가 적절하지 못한 문항이 발견되었다. 그리고 작업기억 측정과제들과 어휘점수 간 상관은 모두 유의하여 공인타당도의 증거를 제시하였다. **논의 및 결론:** 본 연구에서는 활발히 사용되고 있는 작업기억 측정과제의 문항을 체계적으로 검토하는데 있어서 두 검사이론에 입각하여 문항양호도 검증을 시도했다는 데에 의의가 있다. 2차 예비검사의 표본 크기는 작기 때문에 해석상 주의가 요구되며, 추후 표본을 추가하여 타당도를 보완할 필요가 있다.

핵심어: 작업기억, 비단어 따라말하기, 문장 따라말하기, 문항분석, 문항반응이론

이 논문은 2020년 대한민국 과학기술정보통신부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2019R1A2C1007488).

참고문헌

- 김신영, 임동선 (2015). 어휘발달지체 아동을 대상으로 한 작업기억 증제가 작업기억 및 언어능력에 미치는 영향. *Communication Sciences & Disorders*, 20(4), 469-489.
- 김영태 (1997). 한국 2-4 세 아동의 발화길이에 관한 기초연구. *언어청각장애연구*, 2(1), 5-25.
- 김영태, 성태제, 이윤경 (2003). *취학전 아동의 수용언어 및 표현언어 척도(PRES)*. 서울: 서울장애인종합복지관.
- 김영태, 홍경훈, 김경희, 장혜성, 이주연 (2010). *수용·표현어휘력검사(REVT)*. 서울: 장애인종합복지관.
- 문수백, 변창진 (2003). *K-ABC 교육·심리측정도구(Korean-Kaufman assessment battery for children)*. 서울: 학지사.
- 성태제 (1991). *문항반응이론 입문*. 서울: 양서원.
- 성태제 (2002). *현대교육평가*. 서울: 학지사.
- 성태제 (2016). *문항반응이론의 이해와 적용: 측정과 평가 2*. 서울: 교육과학사.
- 양윤희, 임동선 (2018). 학령전기 어휘발달지연 및 포래 아동의 어휘습득을 위한 집행기능의 역할. *Communication Sciences & Disorders*, 23(1), 1-17.
- 양윤희, 임동선, 김신영, 한지윤 (2013). 학령 전 어휘발달지체 및 일반 아동의 비단어 따라 말하기, 빠른 우연학습(Quick Incidental Learning)과 수용 어휘와의 관계. *Communication Sciences & Disorders*, 18(4), 379-391.
- 양윤희, 임동선, 배경란 (2015). 학령기 단순언어장애 아동의 어휘학습 예측요인: 의미점화 및 간섭 효과. *Communication Sciences & Disorders*, 20(1), 1-12.
- 오다연, 임동선(2013). 2-3세 말늦은 아동과 정상 아동의 비단어 따라말하기와 문장따라말하기 수행능력. *Communication Sciences & Disorders*, 18(3), 277-287.

- 윤효진 (2015). 초등 저학년과 고학년의 읽기이해 예측 요인 연구. *Communication Sciences & Disorders*, 20(4), 536-546.
- 이상금, 정세화, 이은화, 이정환 (1972). 3·4·5세 아동의 회화에 나타난 어휘조사. *교육과학연구*, 1, 43-133.
- 이연섭, 권경안, 김성일 (1979). *한국 아동의 구문발달 (I)*. 서울: 한국교육개발원.
- 임동선, 김경태, 양윤희 (2016). 언어장애아동 판별을 위한 구어 및 시공간적 작업기억의 효용성 탐색. *Communication Sciences & Disorders*, 21(2), 193-205.
- 임동선, 양윤희, 김신영 (2015). 단순언어장애 아동과 일반 아동의 작업기억 제시방식 및 과제유형에 따른 수행능력 비교. *Communication Sciences & Disorders*, 20(1), 13-23.
- 한우주, 임동선 (2016). 단순언어장애 아동의 구문접화 방식에 따른 복문 산출 능력과 암묵적 학습 및 작업기억 간의 관계. *Communication Sciences & Disorders*, 21(4), 590-604.
- 황민아, 임종아, 최경순, 고선희, 최소영, 김주형, 전미영 (2016). 주의력결핍 과잉행동장애 아동의 단어재인에서 간섭자극 억제능력. *Communication Sciences & Disorders*, 21(1), 121-130.

ORCID

한지윤(제1저자, 박사과정생 <https://orcid.org/0000-0002-7570-6679>)
임동선(교신저자, 교수 <https://orcid.org/0000-0001-8254-9504>)
석혜은(공동저자, 박사 <https://orcid.org/0000-0003-4273-8655>)